Region-based approximation of probability distributions (for visibility between imprecise points among obstacles)[1]

Kevin Buchin* Irina Kostitsyna* Maarten Löffler† Rodrigo I. Silveira‡

Abstract

Let p and q be two imprecise points, given as probability density functions on $\mathbb{R}^2$, and let $\mathcal{R}$ be a set of n line segments in $\mathbb{R}^2$. We study the problem of approximating the probability that p and q can see each other; that is, that the segment connecting p and q does not cross any segment of $\mathcal{R}$. To solve this problem, we approximate each density function by a weighted set of polygons; a novel approach to dealing with probability density functions in computational geometry.

1 Introduction

Data imprecision is an important obstacle to the application of geometric algorithms to real-world problems. In the computational geometry literature, various models to deal with data imprecision have been suggested. Most generally, in this paper we describe the location of each point by a probability distribution μ_i (for instance by a Gaussian distribution). This model is often not worked with directly because of the computational difficulties arisen from its generality.

These difficulties can often be addressed by approximating the distributions by point sets. For instance, for tracking uncertain objects a particle filter uses a discrete set of locations to model uncertainty [11]. Löffler and Phillips [8] and Jørgenson *et al.* [6] discuss several geometric problems on points with probability distributions, and show how to solve them using discrete point sets (or *indecisive* points) that have guaranteed error bounds. More specifically, a 2-dimensional point set P is an ε-*quantization* of an xy-monotone function F (such as a cumulative probability density function), if for every point q in the plane the fraction of P dominated by q differs from $F(q)$ by at most ε.

*Dept. of Mathematics and Computer Science, TU Eindhoven; {k.a.buchin, i.kostitsyna}@tue.nl.
†Department of Computing and Information Sciences, Utrecht University; m.loffler@uu.nl.
‡Dept. de Matemática & CIDMA, Universidade de Aveiro, and Dept. Matemàtica Aplicada II, Universitat Politècnica de Catalunya; rodrigo.silveira@ua.pt.

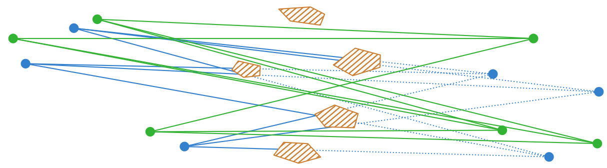


Figure 1: Two pairs of point sets on opposite sides of a collection of obstacles. The green points can all see each other, whereas none of the blue points can.

Imprecise points appear naturally in many applications. They play an important role in databases [1], machine learning [3], and sensor networks [12], where a limited number of probes from a certain data set are gathered, each potentially representing the true location of a data point. Alternatively, data points may be obtained using imprecise measurements or are the result of inexact earlier computations.

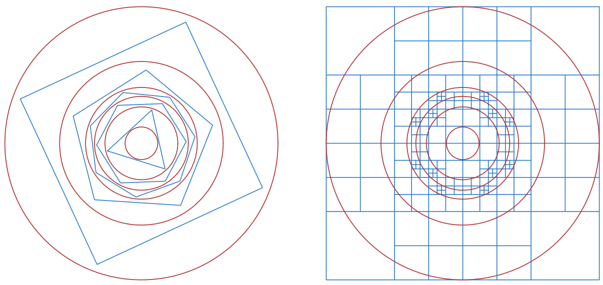
Even though a point set may be a provably good approximation of a probability distribution, this is not good enough in all applications. Consider, for example, a situation where we wish to model visibility between imprecise points among obstacles. When both points are given by a probability distribution, naturally there is a probability that the two points see each other. However, when we discretise the distributions, the choice of points may greatly influence the resulting probability, as illustrated in Figure 1.

Instead, we may approximate distributions by regions. The concept of describing an imprecise point by a region or shape was first introduced by Guibas *et al.* [4], motivated by finite coordinate precision, and later studied extensively in a variety of settings [5, 2, 9, 10, 7].

In this work we show how to use region-based approximation of point distributions to solve algorithmic problems on (general) imprecise points. In Section 2 we discuss several ways to do this. In Section 3, we focus on a geometric problem for which previous point-based methods do not work well: visibility computations between imprecise points.

[1]This is an extended abstract of a presentation given at EuroCG 2014. It has been made public for the benefit of the community and should be considered a preprint rather than a formally reviewed paper. Thus, this work is expected to appear in a conference with formal proceedings and/or in a journal.

This is an extended abstract of a presentation given at EuroCG 2014. It has been made public for the benefit of the community and should be considered a preprint rather than a formally reviewed paper. Thus, this work is expected to appear in a conference with formal proceedings and/or in a journal.



(a) (b)

Figure 2: A Gaussian distribution, given by isolines at ε levels. (a) Approximation by polygons. (b) Approximation by quadtree.

2 Region-based approximation

Assume that an imprecise point p is given by a probability distribution μ. We wish to describe μ by a set of weighted regions $\mathcal{M}$ that provide an additive ε-approximation of the distribution: for any point in the plane, the sum of the weights of the regions containing q differs from $\mu(q)$ by at most ε.

One approach to attack this problem is to consider the *isolines* of the probability density function $f(\cdot)$. These are the curves where $f(\cdot)$ is exactly $k\varepsilon$, for some integer k, and they separate the plane into regions where $f(\cdot)$ has a value between $i\varepsilon$ and $(i+1)\varepsilon$. Note that if we could take the regions formed by the isolines, and give each of them a weight of ε, they would form a valid ε-approximation of f. However, the isolines are not generally polygonal. Instead, we note that if we take any polygons that stay between two consecutive isolines, and use these as polygons with weight ε, they are guaranteed to form a 2ε-approximation. Figure 2(a) illustrates this.

Of course, the complexity of the polygons depends on the specific distribution. In the following we focus on Gaussian distributions, because they are natural and likely to occur in applications.

Theorem 1 *A Gaussian distribution with standard deviation σ can be ε-approximated by $O(\sigma^{-2}\varepsilon^{-1})$ polygons of total complexity $O(\sigma^{-4}\varepsilon^{-2})$.*

Proof. The isolines of a bivariate Gaussian distribution are concentric circles that subdivide $\mathbb{R}^2$ into annuli, and we wish to compute a polygon that stays within each annulus. We observe that the complexity of such a polygon depends only on the relative width of its annulus; that is, given an annulus with inner radius r and outer radius r', we can fit a regular $\lceil \pi / \arccos \frac{r}{r'} \rceil$-gon. Refer to Figure 2(a) for some examples.

The probability density function with standard deviation σ is given by the equation

$$f(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$$

The number of annuli depends on the height of the peak of the function we wish to approximate, which is at $\frac{1}{2\pi\sigma^2}$, so we need $k = \frac{1}{2\pi\sigma^2\varepsilon}$ isolines.

If we solve $f(x,0) = i\varepsilon$ for x, we get

$$x = \sqrt{-2\sigma^2 \log(2\pi\sigma^2 i\varepsilon)}$$

so the ith annulus has relative width

$$\frac{r}{r'} = \sqrt{\frac{\log(2\pi\sigma^2 i\varepsilon)}{\log(2\pi\sigma^2(i+1)\varepsilon)}}.$$

Hence, the total complexity of all polygons is

$$\sum_{i=1}^{k} \left\lceil \pi / \arccos \sqrt{\frac{\log(2\pi\sigma^2 i\varepsilon)}{\log(2\pi\sigma^2(i+1)\varepsilon)}} \right\rceil,$$

which we rather coarsely bound by k times the maximum of these terms, attained at $i = k/2$. We obtain:

$$\frac{1}{2\pi\sigma^2\varepsilon} \left\lceil \pi / \arccos \sqrt{\frac{\log 1/2}{\log(1/2 + 2\pi\sigma^2\varepsilon)}} \right\rceil.$$

As the argument of the arccos approaches 1, the value approaches 0 as the square of the argument, leading to a $O(\frac{1}{\sigma^2\varepsilon})$ growth rate. The lemma follows. $\square$

Alternatively, we may subdivide space into grid cells and give each cell a weight depending on the value of f. The advantage of a grid-based approach is that the subdivision of the plane does not depend on the actual distributions, and that squares are particularly nice polygons. A problem with this approach is that the resolution of the grid depends on the steepest part of f: when the value of f varies by more than ε in a cell, the approximation is not valid. Instead, we may also choose to compute a non-uniform grid, for example based on a quadtree. If we use a quadtree to subdivide $\mathbb{R}^2$ until no cell is crossed by more than one isoline, and we weigh a cell crossed by the ith isoline by $i\varepsilon$, we again obtain a 2ε-approximation. Figure 2(b) illustrates this.

3 Visibility between two regions

Consider a set of obstacles $\mathcal{R}$ in the plane. We assume that the obstacles are disjoint simple convex polygons with m vertices in total. For two imprecise points with probability distributions μ_1 and μ_2 we approximate them with two sets of weighted regions $\mathcal{M}_1$ and $\mathcal{M}_2$,

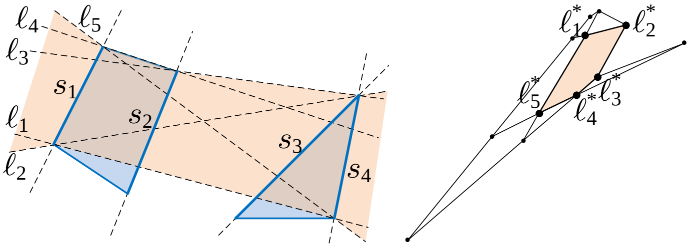


Figure 3: Left: Two polygons P_1 and P_2 in primary space. The orange region is the set of lines intersecting P_1 and P_2 through s_1, s_2, s_3, s_4. Right: Partition L^* in dual space. The orange cell corresponds to all lines in the primary space intersecting the same four segments s_1, s_2, s_3, s_4.

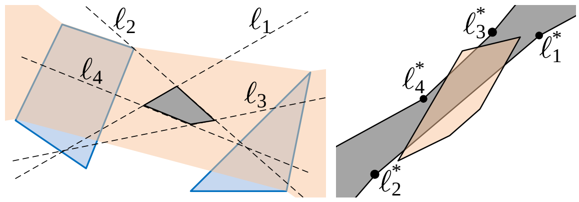


Figure 4: Left: Polygons P_1 and P_2 and an obstacle in between in the primary space. Right: The "hour-glass" shape H^* in the dual space that corresponds to a set H of all lines in the primary space that intersect the obstacle.

each consisting of convex polygons. For every pair of polygons $P_1 \subset \mathcal{M}_1$ and $P_2 \subset \mathcal{M}_2$, we compute the probability that a point p_1 chosen uniformly at random from P_1 can see a point p_2 chosen uniformly at random from P_2. We say that two points can "see" each other if and only if the straight line segment connecting them does not intersect any obstacle from $\mathcal{R}$. The probability of two points $p_1 = (x_1, y_1) \in P_1$ and $p_2 = (x_2, y_2) \in P_2$ seeing each other can be computed by the formula:

$$prob = \frac{\iiiint v(x_1, y_1, x_2, y_2) \mathrm{d}x_1 \mathrm{d}y_1 \mathrm{d}x_2 \mathrm{d}y_2}{\iiiint \mathrm{d}x_1 \mathrm{d}y_1 \mathrm{d}x_2 \mathrm{d}y_2}, \qquad (1)$$

where $v(x_1, y_1, x_2, y_2)$ is 1 if the points see each other, and 0 otherwise.

To compute *prob* we consider a dual space where a point with coordinates (α, β) corresponds to a line $y = \alpha x - \beta$ in the primary space. We construct a region L^* in the dual space that corresponds to the set L of lines that stab both P_1 and P_2. This region can be partitioned into cells, each corresponding to a set of lines that cross the same four segments of P_1 and P_2 (refer to Figure 3). The following follows from the fact that each vertex of L^* corresponds to a line in primary space through two vertices of P_1 and P_2.

Lemma 2 *Given two convex polygons P_1 and P_2 of total size n, the complexity of partition L^* in the dual space that corresponds to a set of lines L that stab P_1 and P_2 is $O(n^2)$.*

For each obstacle $h \subset \mathcal{R}$ we construct a region H^* in the dual space, that corresponds to the set of lines that intersect h. H^* has an "hour-glass" shape (refer to Figure 4). We now compute the subdivision Ł of the dual plane resulting from overlaying the partition L^* and the regions H^*. Since the objects involved are bounded by a total of $O(m + n)$ line segments in the primal space, Ł has complexity $O((m + n)^2)$.

First consider the case that P_1, P_2 and the obstacles are disjoint. We can assume that all obstacles lie in the convex hull of P_1 and P_2. Then a pair of points from P_1 and P_2 see each other exactly if the line through the points does not intersect an obstacle. Thus, we only need to identify the cells in Ł not intersecting any of the regions H^*, and integrate over these cells. Details on evaluating the integral for one cell is given in Section 4. Overall this case can be handled in $O((m+n)^2)$ time. Next, consider the case that P_1 and P_2 are disjoint but might intersect obstacles. Now we need to consider the length of each line segment from the last obstacle in P_1 to the boundary and from the boundary of P_2 to the first obstacle. We can annotate the cells of Ł with this information by a traversal of Ł. Between neighboring cells this information can be updated in constant time. Thus, this case can be handled with the same asymptotic running time as the previous case. As a third case, consider P_1 overlapping P_2 but with no obstacles in the overlap area. The computations needed remain the same as in the case of non-overlapping P_1 and P_2.

Finally, we consider the general case, in which obstacles might also lie in the overlap of P_1 and P_2. In the cells of Ł that correspond to the overlap of P_1 and P_2 we now need to consider the sum of the lengths of each line segment between boundaries of obstacles. Again we traverse Ł maintaining the ordered list of intersected obstacle boundaries. Within a cell we use this to compute the sum of lengths in $O(m)$ time.

Lemma 3 *Given two polygons P_1 and P_2 of total size n and obstacles of total complexity m, we can compute the probability that a pair of points drawn uniformly at random from $P_1 \times P_2$ can see each other in $O(m(m+n)^2)$ time, assuming we can compute the necessary information within each cell.*

4 Computing the probability for a single cell

For simplicity of presentation, we assume that P_1 and P_2 are separable by a vertical line, and P_1 and P_2 are disjoint from $\mathcal{R}$. This will allow us to write the solution in a more concise way without loss of generality.

Consider line ℓ, given by the formula $y = \alpha x - \beta$, that goes through two points $p_1(x_1, y_1) \in P_1$ and $p_2(x_2, y_2) \in P_2$. In the dual space, point ℓ^*, corresponding to line ℓ, has coordinates (α, β). Substitute variables y_1 and y_2 in Formula 1 with α and β: $(x_1, y_1, x_2, y_2) \leftarrow (x_1, \alpha, x_2, \beta)$, where $\alpha(x_1, y_1, x_2, y_2) = y_2 - y_1/x_2 - x_1$ and $\beta(x_1, y_1, x_2, y_2) = (x_1 y_2 - x_2 y_1)/(x_2 - x_1)$. We can express the probability of two points, distributed uniformly at random in P_1 and P_2, seeing each other as

$$prob = \frac{\iiiint v(x_1, \alpha, x_2, \beta)|J| \mathrm{d}x_1 \mathrm{d}x_2 \mathrm{d}\alpha \mathrm{d}\beta}{\iiiint |J| \mathrm{d}x_1 \mathrm{d}x_2 \mathrm{d}\alpha \mathrm{d}\beta}, \qquad (2)$$

where

$$J = \det \begin{bmatrix} \frac{dy_1}{\mathrm{d}\alpha} & \frac{\mathrm{d}y_1}{\mathrm{d}\beta} \\ \frac{dy_2}{\mathrm{d}\alpha} & \frac{\mathrm{d}y_2}{\mathrm{d}\beta} \end{bmatrix} = \frac{1}{\det \begin{bmatrix} \frac{\mathrm{d}\alpha}{\mathrm{d}y_1} & \frac{\mathrm{d}\beta}{\mathrm{d}y_1} \\ \frac{\mathrm{d}\alpha}{\mathrm{d}y_2} & \frac{\mathrm{d}\beta}{\mathrm{d}y_2} \end{bmatrix}} = x_2 - x_1 \,.$$

The denominator of (2) can be written as a sum of integrals over all cells of partition L^* in the dual space:

$$\sum_{C \subset L^*} \iint_C \left(\int_{X_1(\alpha,\beta)}^{X_2(\alpha,\beta)} \int_{X_3(\alpha,\beta)}^{X_4(\alpha,\beta)} (x_2 - x_1) \mathrm{d}x_2 \mathrm{d}x_1 \right) \mathrm{d}\alpha \mathrm{d}\beta \,,$$

where $X_1(\alpha, \beta)$, $X_2(\alpha, \beta)$, $X_3(\alpha, \beta)$, and $X_4(\alpha, \beta)$ are the x-coordinates of intersections of line $y = \alpha x - \beta$ with the boundary segments of P_1 and P_2.

The numerator of (2) can be written as a sum of integrals over all cells of partition $L^* \setminus \cup_h H^*$ in the dual:

$$\sum_{C \subset L^* \setminus \cup_h H^*} \iint_C \left(\int_{X_1(\alpha,\beta)}^{X_2(\alpha,\beta)} \int_{X_3(\alpha,\beta)}^{X_4(\alpha,\beta)} (x_2 - x_1) \mathrm{d}x_2 \mathrm{d}x_1 \right) \mathrm{d}\alpha \mathrm{d}\beta.$$

For more details on calculating the integrals we refer the reader to the full version of this article.

Theorem 4 *Given two convex polygons P_1 and P_2 of total size n and a set of obstacles of total size m, we can compute the probability that a point p_1 chosen uniformly at random in P_1 sees a point p_2 chosen uniformly at random in P_2 in $O(m(m + n)^2)$ time.*

5 Main result

Combining Theorems 1 and 4, our main result follows:

Theorem 5 *Given two imprecise points, modelled as Gaussian distributions μ_1 and μ_2 with standard deviations σ_1 and σ_2, and n obstacles, we can ε-approximate the probability that p and q see each other in $O(\sigma_1^{-2}\sigma_2^{-2}\varepsilon^{-2}(\sigma_1^{-2} + \sigma_2^{-2})\varepsilon^{-1}((\sigma_1^{-2} + \sigma_2^{-2})\varepsilon^{-1} + n)^2)$ time.*

Proof. According to Theorem 1, we need to solve $O(\sigma_1^{-2}\sigma_2^{-2}\varepsilon^{-2})$ individual problems. For each, we have $m = O((\sigma_1^{-2} + \sigma_2^{-2})\varepsilon^{-1})$, so using Theorem 4 we solve them in $O((\sigma_1^{-2} + \sigma_2^{-2})\varepsilon^{-1}((\sigma_1^{-2} + \sigma_2^{-2})\varepsilon^{-1} + n)^2)$ time. This leads to $O(\sigma_1^{-2}\sigma_2^{-2}\varepsilon^{-2}(\sigma_1^{-2} + \sigma_2^{-2})\varepsilon^{-1}((\sigma_1^{-2} + \sigma_2^{-2})\varepsilon^{-1} + n)^2)$ running time . $\square$

Acknowledgments. K.B., I.K., and M.L. are supported by the Netherlands Organisation for Scientific Research (NWO) under grant no. 612.001.106, 612.001.207, and 639.021.123, respectively. R.S. was funded by Portuguese funds through CIDMA and FCT, within project PEst-OE/MAT/UI4106/2014, and by FCT grant SFRH/BPD/88455/2012. In addition, R.S. was partially supported by projects MINECO MTM2012-30951/FEDER, Gen. Cat. DGR2009SGR1040, and by ESF EUROCORES program EuroGIGA-ComPoSe IP04-MICINN project EUI-EURC-2011-4306.

References

[1] P. K. Agarwal, S.-W. Cheng, Y. Tao, and K. Yi. Indexing uncertain data. In *PODS*, pages 137–146, 2009.

[2] D. Bandyopadhyay and J. Snoeyink. Almost-Delaunay simplices: Nearest neighbor relations for imprecise points. In *SODA*, pages 410–419, 2004.

[3] J. Bi and T. Zhang. Support vector classification with input data uncertainty. In *NIPS*, 2004.

[4] L. J. Guibas, D. Salesin, and J. Stolfi. Epsilon geometry: building robust algorithms from imprecise computations. In *SoCG*, pages 208–217, 1989.

[5] L. J. Guibas, D. Salesin, and J. Stolfi. Constructing strongly convex approximate hulls with inaccurate primitives. *Algorithmica*, 9:534–560, 1993.

[6] A. Jørgensen, M. Löffler, and J. Phillips. Geometric computations on indecisive points. In *WADS*, pages 536–547, 2011.

[7] M. Löffler. *Data Imprecision in Computational Geometry.* PhD thesis, Utrecht University, 2009.

[8] M. Löffler and J. Phillips. Shape fitting on point sets with probability distributions. In *ESA*, pages 313–324, 2009.

[9] T. Nagai and N. Tokura. Tight error bounds of geometric problems on convex objects with imprecise coordinates. In *Jap. Conf. on Discrete and Comput. Geom.*, pages 252–263, 2000.

[10] Y. Ostrovsky-Berman and L. Joskowicz. Uncertainty envelopes. In *EuroCG*, pages 175–178, 2005.

[11] R. van der Merwe, A. Doucet, N. de Freitas, and E. Wan. The unscented particle filter. In *Adv. Neural Inf. Process. Syst.*, volume 8, pages 351–357, 2000.

[12] Y. Zou and K. Chakrabarty. Uncertainty-aware and coverage-oriented deployment of sensor networks. *J. Parallel Distrib. Comput*, pages 788–798, 2004.